

Statistical Inference of In Vivo Properties of Human DNA Methyltransferases from Double-Stranded Methylation Patterns

Audrey Q. Fu^{1*†}, Diane P. Genereux², Reinhard Stöger³, Alice F. Burden², Charles D. Laird², Matthew Stephens⁴

1 Department of Physiology, Development and Neuroscience, Cambridge Systems Biology Centre, University of Cambridge, Cambridge, United Kingdom, **2** Department of Biology, University of Washington, Seattle, Washington, United States of America, **3** School of Biosciences, University of Nottingham, Sutton Bonington Campus, Leicestershire, United Kingdom, **4** Departments of Human Genetics and Statistics, University of Chicago, Chicago, Illinois, United States of America

Abstract

DNA methyltransferases establish methylation patterns in cells and transmit these patterns over cell generations, thereby influencing each cell's epigenetic states. Three primary DNA methyltransferases have been identified in mammals: DNMT1, DNMT3A and DNMT3B. Extensive in vitro studies have investigated key properties of these enzymes, namely their substrate specificity and processivity. Here we study these properties in vivo, by applying novel statistical analysis methods to double-stranded DNA methylation patterns collected using hairpin-bisulfite PCR. Our analysis fits a novel Hidden Markov Model (HMM) to the observed data, allowing for potential bisulfite conversion errors, and yields statistical estimates of parameters that quantify enzyme processivity and substrate specificity. We apply this model to methylation patterns established in vivo at three loci in humans: two densely methylated inactive X (Xi)-linked loci (*FMR1* and *G6PD*), and an autosomal locus (*LEP*), where methylation densities are tissue-specific but moderate. We find strong evidence for a high level of processivity of DNMT1 at *FMR1* and *G6PD*, with the mean association tract length being a few hundred base pairs. Regardless of tissue types, methylation patterns at *LEP* are dominated by DNMT1 maintenance events, similar to the two Xi-linked loci, but are insufficiently informative regarding processivity to draw any conclusions about processivity at that locus. At all three loci we find that DNMT1 shows a strong preference for adding methyl groups to hemi-methylated CpG sites over unmethylated sites. The data at all three loci also suggest low (possibly 0) association of the de novo methyltransferases, the DNMT3s, and are consequently uninformative about processivity or preference of these enzymes. We also extend our HMM to reanalyze published data on mouse DNMT1 activities in vitro. The results suggest shorter association tracts (and hence weaker processivity), and much longer non-association tracts than human DNMT1 in vivo.

Citation: Fu AQ, Genereux DP, Stöger R, Burden AF, Laird CD, et al. (2012) Statistical Inference of In Vivo Properties of Human DNA Methyltransferases from Double-Stranded Methylation Patterns. PLoS ONE 7(3): e32225. doi:10.1371/journal.pone.0032225

Editor: Pierre-Antoine Defossez, Université Paris-Diderot, France

Received: May 23, 2011; **Accepted:** January 24, 2012; **Published:** March 19, 2012

Copyright: © 2012 Fu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the National Institutes of Health Grants HD002274 and GM077464 to C.D.L. (<http://www.nih.gov/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: audreyqfu@uchicago.edu

† Current address: Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

Introduction

DNA methyltransferases establish methylation patterns in cells and transmit these patterns over cell generations, thereby influencing each cell's epigenetic states. (See [1] for an overview of methyltransferases, and Supplementary Material of [2] for an introduction to DNA methylation aimed at non-biologists.) Three primary DNA methyltransferases have been identified in mammals: DNMT1, DNMT3A and DNMT3B [3,4]. Whereas the DNMT3s are mostly responsible for establishing methylation patterns during early development and are therefore commonly known as the de novo methyltransferases, DNMT1 is mostly responsible for maintaining existing methylation patterns over somatic cell divisions, and is therefore commonly known as the maintenance methyltransferase [1].

A central component of the widely accepted model for the maintenance of DNA methylation in eukaryotes is processive actions

of the maintenance methyltransferase DNMT1 at hemimethylated CpG dyads shortly after DNA replication (Figure 1; [1]). This model relies on two properties of DNMT1: substrate specificity (i.e., acting in different ways or at different rates on different types of substrate) and processivity (i.e., associating consecutively at multiple sites along the DNA). These are key properties of DNA methyltransferases and many other DNA-binding enzymes [1,5], and both properties have been investigated extensively in vitro.

Regarding substrate specificity, in vitro experiments show that DNMT1 preferentially adds methyl groups to the cytosines in daughter-strand CpGs that pair with methylated, rather than unmethylated, parent-strand CpGs (i.e., hemimethylated CpG dyads), thus maintaining methylation at these CpG sites over cell generations [6]. Such a preference for hemimethylated CpG dyads was predicted for maintenance methyltransferases as early as 1968 [7], and is now commonly measured in terms of the “hemi-preference ratio”. This ratio represents the relative rates with

March 2012 | Volume 7 | Issue 3 | e32225

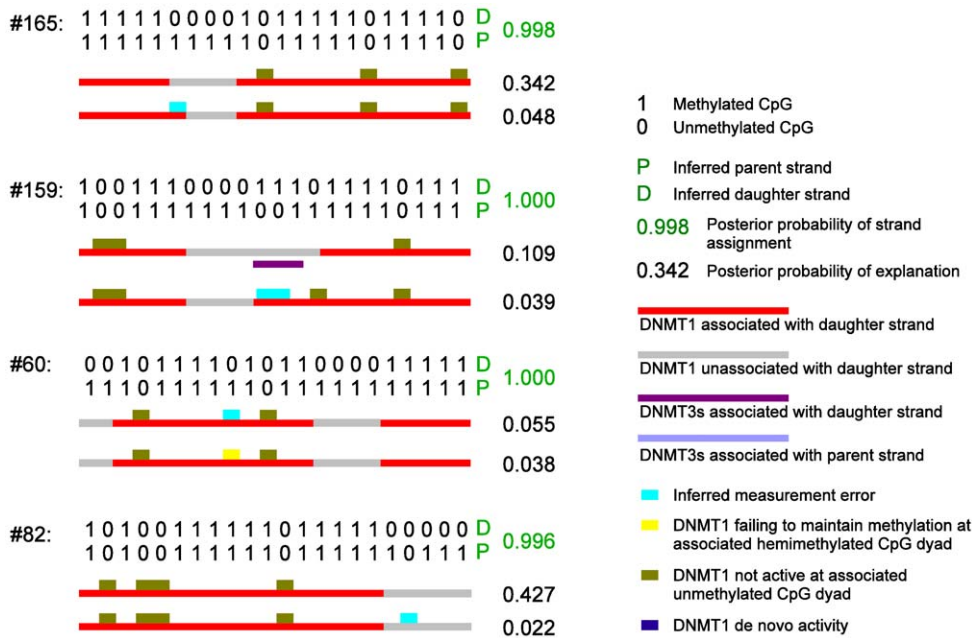


Figure 2. Four most informative *FMR1* methylation patterns and inference for each of them under HMM. Each pattern contains a pair of parent and daughter strands; which is the parent strand and which is the daughter strand is not known directly. Our HMM infers probabilistically the strand assignments for each pattern. These posterior probabilities are shown in green for the indicated parent and daughter strand assignment (indicated by letters P and D in green). These patterns from our *FMR1* data contain long runs of hemimethylated CpG dyads with their methyl groups present on the same strand. The top two most likely explanations, among all 10^{39} possible explanations, are shown for each pattern. Symbols indicate possible states and activities of the methyltransferases. Not all symbols are used here. The effect of these four patterns on parameter estimation is further shown in Supplementary Figure 13 in Materials S1.
doi:10.1371/journal.pone.0032225.g002

and the DNMT3s) were either associated or not associated with the DNA at a CpG site. We model these association/non-association states as a Markov process along the DNA. The model for the observed data based on these unobserved (“hidden”) states is then a Hidden Markov Model (HMM [24]). For in vivo data, we cannot rule out that other (perhaps unidentified) enzymes than DNMT1 and the DNMT3s may have also contributed to the observed methylation patterns. To allow for this, our references to DNMT1 could be interpreted broadly as referring to the enzymes whose activities are primarily maintenance methylation, and references to DNMT3s could be interpreted broadly as enzymes whose activities are primarily de novo methylation.

More specifically, the hidden Markov process in the HMM can be decomposed into three independent Markov processes: the first representing association of DNMT1, which we assume to act only on the daughter strand [10], the second representing the association of the DNMT3s on the parent strand, and the third representing the association of the DNMT3s on the daughter strand. We use the subscripts $1d$, $3p$ and $3d$ to refer to each of these processes. We characterize each Markov process by two transition probabilities: the reassociating probability per bp, τ , that unassociated molecules of the methyltransferase become associated with DNA over 1 bp, and the dissociating probability per bp, ρ , that associated molecules of the methyltransferase become unassociated from DNA over 1 bp. The transition probability matrix over 1 bp for each Markov process can then be written as in Eq. (1).

$$\mathbf{G}_1 = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 1-\tau & \tau \\ \rho(1-\tau) & (1-\rho)+\rho\tau \end{pmatrix} \end{matrix} \quad (1)$$

The transition probability matrix between two CpG sites k bp apart is then $\mathbf{G}_k = \mathbf{G}_1^k$.

Under this parameterisation, two events contribute to the probability of a methyltransferase staying associated at two consecutive sites 1 bp apart (i.e., the bottom right entry in \mathbf{G}_1): (i) processivity, whereby the methyltransferase stays associated from site 1 to site 2 without ever dissociating from DNA. This happens with probability $1-\rho$; and (ii) dissociation-reassociation events, whereby the methyltransferase is associated at site 1, then “falls off”, but becomes re-associated with DNA at site 2. This happens with probability $\rho\tau$. We quantify the strength of processivity, which corresponds to event (i), by the dissociating probability ρ : the value $\rho=1$ corresponds to “no processivity”, where the methyltransferase dissociates from DNA at every base pair and the association state is independent at each site (i.e., the two rows of \mathbf{G}_1 are identical); values of $\rho<1$ correspond to processive behaviour, and the smaller the value of ρ , the stronger the processivity. This parameter translates into the expected length of each association tract of $1/\rho$ bp, which is a more conventional measure of processivity and quantifies processivity directly in terms of tract length. Because ρ concerns only enzyme molecules that are already associated, this mean length of association tracts excludes multiple dissociation-reassociation events at consecutive CpG sites, which could be mistaken for processivity. Similarly, the expected length of non-association tracts, which are gaps between association tracts, has mean $1/\tau$ bp. Unlike ρ , the reassociating probability τ could be driven principally by concentrations of unassociated methyltransferase molecules in the nucleus.

Parameters ρ and τ together determine the average frequency with which the methyltransferase is associated with the DNA at each CpG site. Using π to denote this frequency, we have

$$\pi = \tau / (\tau + \rho(1 - \tau)). \quad (2)$$

In addition to these parameters relating to processivity, our model also has parameters for the methylation activities of the methyltransferases when they are associated with DNA, and parameters for the measurement errors that can result from bisulfite conversion (see Materials and Methods for detail). Together these parameters, the “emission probabilities” of the HMM, describe how association or non-association states of methyltransferases give rise to observed methylation states on the pairs of parent and daughter strands, subject to measurement errors (see Materials and Methods for detail). We allow the methyltransferases to methylate daughter CpGs at associated hemi-methylated sites with probability μ_{1d} for DNMT1 (or μ_{3d} for the DNMT3s) and at associated unmethylated sites with probability δ_{1d} for DNMT1 (or δ_{3d} for the DNMT3s). The ratios μ_{1d}/δ_{1d} and μ_{3d}/δ_{3d} are termed the “hemi-preference ratios” for DNMT1 and the DNMT3s, respectively. For the parent strand, we make the simpler assumption that the DNMT3s always add a methyl group to the associated CpG.

The above model is very general, allowing for a complex combination of behaviors of the methyltransferases. In applications, it can be helpful to constrain this model in various ways, either to deal with data collected from particular experimental conditions, or to make parameters more identifiable. For example, imposing the constraint $\mu_{3d} = \delta_{3d} = 1$ yields a more parsimonious model in which the DNMT3s always methylate the associated daughter-strand CpG. This is the model we use for most analyses presented here. However, to attempt to estimate the hemi-preference ratio for the DNMT3s, we impose a different constraint $\tau_{1d} > 0.05$ and $\tau_{3d} < 0.05$ onto the general model. This constraint reflects the setting where DNMT1 is the primary maintenance enzyme, and helps to distinguish the DNMT1 process and the daughter-strand process of the DNMT3s, which would have been indistinguishable otherwise (see Materials and Methods for detail). To analyze the in vitro data on DNMT1 under the same model, we estimate parameters only associated with the DNMT1 process, and fix the reassociating probabilities ρ_{3p} and ρ_{3d} to be 0 and the dissociating probabilities τ_{3p} and τ_{3d} to be 1; these constraints reflect the in vitro setting where the DNMT3s are absent.

We fit the HMM to the data in a Bayesian inference framework [25], using Markov chain Monte Carlo (MCMC) to produce samples from the joint posterior distribution of all parameters in the model given the data (see Materials S1 for details, including specification of relevant prior distributions). At the core of our implementation is the standard forward-backward algorithm in each MCMC iteration for computing the joint likelihood of the parameters given observed methylation patterns. The computational complexity of the forward-backward algorithm for all N patterns across S CpG sites in each MCMC iteration is $O(8^2NS)$, where $8 (=2 \times 2 \times 2)$ is the number of hidden states at each site, with 2 being the two states (associated and unassociated) of each Markov process. We summarize the posterior distributions of the parameters from the Bayesian inference by the posterior median and 80% credible intervals (80% CIs; 10- and 90-percentiles); 80% intervals were used, rather than more conventional 95% intervals, to reduce the impact of the heavy tails of some distributions. This inference procedure accounts for the uncertainty in the data regarding which enzymatic activities produced each observed double-stranded pattern by using a dynamic programming algorithm to sum over all possibilities, weighting each possibility by its probability (Figure 2; see also Materials S1).

Runs of hemi-methylated dyads provide information on processivity

As mentioned above, both processivity and multiple dissociation-reassociation events may give rise to runs of fully methylated dyads. When the observed patterns contain only runs of fully or un-methylated dyads, we cannot tell these two mechanisms apart. Presence of hemimethylated dyads (“hemis” hereinafter) provides additional information. Whereas randomly-distributed hemis in the data suggest that the methyltransferases may have been associated with or dissociated from the dyads randomly, clustered hemis suggest nonrandom enzyme activities. In particular, runs of hemis of the same orientation (i.e., methylated CpGs appearing on the same strand), if observed more often than expected by chance, provide evidence for processivity. For example, the in vitro data on mouse DNMT1 from Goyal et al. [11] (their Fig. 3A) shows multiple very long runs of hemi-methylated sites of the same orientation. Some of these runs contain as many as 20 hemimethylated dyads, with the parent strands being methylated prior to the reaction with DNMT1. These long runs are extremely unlikely if DNMT1 were to associate with the DNA independently at each site. In our in vivo data, similar runs of hemis, although much shorter, also suggest the presence of processivity: indeed a permutation test based on correlations at adjacent CpG sites produces a p value of 0.002, suggesting that these runs of hemis occur more often than expected by chance; see Supplementary Figure 1 in Materials S1. Observations like this motivated the HMM described above, and results from the HMM confirm that these runs of hemi-methylated sites are likely due to DNMT1 being unassociated with the DNA for several sites in succession (see Figure 2 for the top two explanations our HMM inferred for four patterns collected at the *FMRI* locus).

Strong processivity and high hemi-preference ratio of in vivo human DNMT1 at two inactive X (Xi)-linked loci

The methylation patterns at the *FMRI* and *G6PD* loci share many similarities: 77–82% of the CpG sites are fully methylated, whereas only 6–9% are hemimethylated, some of which form a few relatively long runs of hemimethylated dyads (Table 1). Analyses of these two data sets under our HMM also produce similar estimates of the key parameters.

We find that the methylation patterns at both of these Xi-linked loci provide strong evidence for substantial processivity of DNMT1: the estimated dissociating probability for the DNMT1 process, ρ_{1d} , is concentrated on small values near 0 (Table 2; Figure 4(A1)). To investigate the robustness of the estimates, we further ran our HMM on the *FMRI* data with different prior distributions assigned to ρ . The estimates were similar across these runs (Supplementary Figure 2 in Materials S1). These estimates of ρ_{1d} at these two loci imply a mean association tract length of around 600 bp, which is equivalent to about 90 CpG sites (note that there is considerable uncertainty in these estimates, and the 80% CIs span almost 200–2000 bp; Table 2). This inferred length is much greater than the genomic regions covered by our data (142 bp at *FMRI* and 122 bp at *G6PD*; Table 1), as it reflects the expected length of DNMT1 association, had we measured methylation over a much longer genomic region (Figures 2 and 3). Interpreting estimated processivity in terms of tract length allows us to compare our estimates directly with other estimates reported in the literature. On the other hand, we estimate at both loci with strong evidence that the reassociating probability τ_{1d} is not high, with the median being 0.12 at *FMRI* and 0.07 at *G6PD* and the 80% CIs being narrow (Table 2; also see Supplementary Figures 3 and 4 in Materials S1). This estimate is also robust to

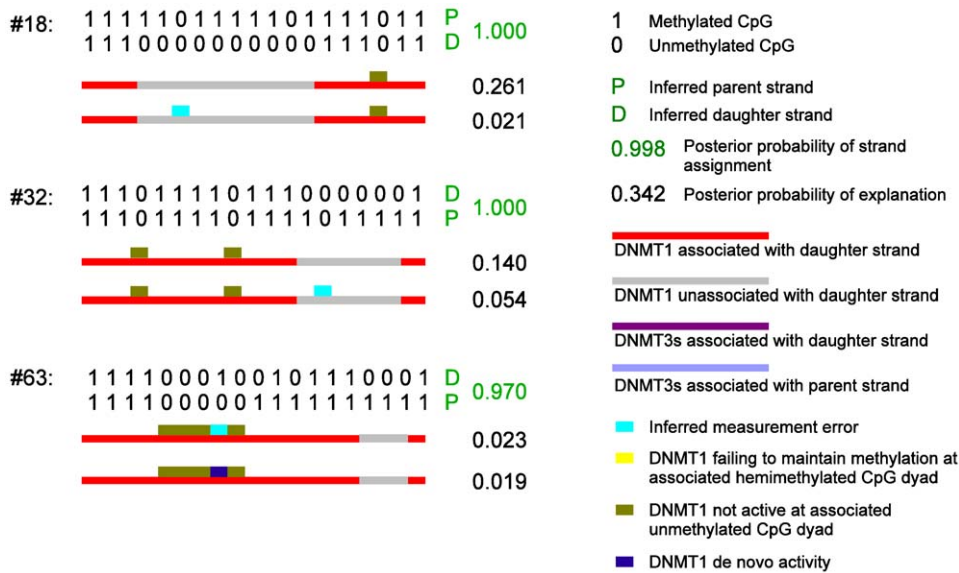


Figure 3. Three most informative *G6PD* methylation patterns and inference for each of them under HMM. Symbols are as in Figure 2. doi:10.1371/journal.pone.0032225.g003

different priors (Supplementary Figure 5 in Materials S1). The high association tract length and low reassociating probability for these two hypermethylated loci imply that strong processivity, rather than random association, of DNMT1 accounts for most of the runs of fully methylated CpG dyads.

Recall that our HMM also allows that DNMT1, when associated with the DNA, adds a methyl group to the daughter strand with probability μ_{1d} if the parent strand is methylated, and with probability δ_{1d} if the parent strand is unmethylated. We estimate μ_{1d} to be close to 1 at both loci (Table 2) and δ_{1d} to be just a few percent, with the median being 0.02 at *FMRI* and 0.06 at *G6PD* (Table 2), consistent with a very low (not excluding 0) level of de novo activity by DNMT1. These estimates indicate that DNMT1 acting in vivo has a strong preference for hemimethylated CpG dyads over unmethylated dyads. Our estimates

of the hemi-preference ratio μ_{1d}/δ_{1d} (58 for *FMRI* and 15 for *G6PD*) fall within the aforementioned range of estimates obtained in vitro (2–200-fold), but the posterior distributions exclude the lower end of this range (Supplementary Figures 3 and 4 in Materials S1). Note that, because the data do not exclude very small values for δ_{1d} , they also cannot exclude extremely large values for the hemi-preference ratio. For this reason Table 2 gives 80% lower confidence bounds, but not upper bounds for this quantity.

Low association level of in vivo human DNMT3s at two Xi-linked loci

At both Xi-linked loci, we estimated that the average frequency of association (π_{3p} and π_{3d}) of the DNMT3s is at most a few

Table 1. Features of the three loci and summary statistics of the methylation patterns in the four human in vivo data sets.

	<i>FMRI</i>	<i>G6PD</i>	<i>LEP</i>	
			In Fat	In Blood
Genomic location	ChrX: 146,800,867-1,008	ChrX: 153,775,537-698	Chr7: 127,881,204-375	
Region length (bp)	142	122	172	
No. of CpG sites	22	19	21	
Median distance (bp) between CpG sites	6.7	6.5	7.5	
No. of ds patterns	169	75	80	34
% of M, H, U*	(82, 6, 12)	(77, 9, 14)	(16, 4, 80)	(40, 5, 55)
Runs of 2 hemis**	11	4	6	1
Runs of 3 hemis	1	1	1	0
Runs of 4 hemis	3	1	0	0
Runs of 5 hemis	0	1	0	0
Runs of >5 hemis	0	0	0	0

*Percentages of methylated, hemimethylated and unmethylated dyads.

**Consecutive hemimethylated CpG dyads with methylated groups appearing on the same strand.

doi:10.1371/journal.pone.0032225.t001

Table 2. Parameter estimates under our HMM for the four human in vivo data sets.

			LEP (Chr 7)	
Parameter	FMR1 $\sqrt{2}$ (Chr X)	G6PD (Chr X)	In Fat	In Blood
Mean association				
length (bp):				
of DNMT1	597 (219–1521)	673 (238–1874)	9 (2–88)*	9 (2–89)*
of DNMT3s on parent	2 (1–6)*	1 (1–3)*	5 (1–17)*	1 (1–4)*
of DNMT3s on daughter	4 (1–14)*	1 (1–3)*	2 (1–14)*	2 (1–3)*
Dissociating				
probability:				
ρ_{1d}	0.002 (0.000–0.005)	0.001 (0.001–0.004)	0.11 (0.01–0.53)*	0.11 (0.01–0.63)*
ρ_{3p}	0.65 (0.16–0.93)*	0.71 (0.35–0.94)*	0.20 (0.06–0.82)*	0.68 (0.28–0.92)*
ρ_{3d}	0.28 (0.07–0.79)*	0.72 (0.38–0.95)*	0.49 (0.07–0.93)*	0.66 (0.29–0.92)*
Reassociating				
probability:				
τ_{1d}	0.12 (0.07–0.21)	0.07 (0.03–0.12)	0.81 (0.21–0.96)*	0.91 (0.60–0.98)*
τ_{3p}	0.01 (0.00–0.03)	0.03 (0.00–0.08)	0.00 (0.00–0.00)	0.00 (0.00–0.01)
τ_{3d}	0.01 (0.00–0.03)	0.03 (0.00–0.10)	0.00 (0.00–0.00)	0.01 (0.00–0.02)
DNMT1:				
Maintenance probability	0.99 (0.97–1.00)	0.98 (0.96–1.00)	0.95 (0.91–0.99)	0.98 (0.95–1.00)
De novo probability	0.02 (0.00–0.05)	0.06 (0.01–0.14)	0.00 (0.00–0.01)	0.01 (0.00–0.02)
Hemi-preference ratio	58 (> 25)	15 (> 8)	628 (> 208)	94 (> 50)
Average association level				
π_{1d}	0.99 (0.98–1.00)	0.98 (0.96–0.99)	0.96 (0.89–1.00)	0.99 (0.96–1.00)
π_{3p}	0.02 (0.00–0.05)	0.04 (0.00–0.10)	0.01 (0.00–0.02)	0.01 (0.00–0.02)
π_{3d}	0.05 (0.01–0.08)	0.05 (0.01–0.12)	0.00 (0.00–0.02)	0.01 (0.00–0.03)
Measurement error				
probability (c)	0.02 (0.01–0.03)	0.02 (0.01–0.03)	0.02 (0.01–0.05)	0.02 (0.01–0.04)

Medians and 80% credible intervals (CIs) of the posterior distribution of the parameters under our HMM are reported. The lower and upper limit of the 80% CI represent the 10- and 90-percentile, respectively. For hemi-preference ratios, one-sided 80% CIs are reported; that is, the lower limit is the 20-percentile of the distribution. Measurement error probability c refers specifically to the inappropriate bisulfite conversion probability per CpG per strand (see Materials and Methods for detail). Except for mean association lengths and hemi-preference ratios, estimates reported here were obtained under a uniform (0,1) prior. Entries with an * are sensitive to the choice of the prior distribution; in other words, the data are less informative about these parameters.

doi:10.1371/journal.pone.0032225.t002

percent, if non-zero, on either the post-replication parent strand or the daughter strand (Table 2). Such a low average level of association is consistent with a low reassociating probability τ for the DNMT3s: the estimated reassociating probability of the DNMT3s is also not much different from 0 on either strand (Table 2). Because only a limited number of observed methylation events could be attributed to the DNMT3s, there is not much information regarding their processivity or hemi-preference ratio. The posterior distributions of the dissociating probabilities ρ_{3p} and ρ_{3d} of the DNMT3s are not much different from the uniform prior distribution we assumed (Figure 4(A2)–(A3)), indicating that the data were not informative enough to alter this prior. Thus, these results suggest that the DNMT3s were not very active during the process the observed methylation patterns were formed, and that our data at the two Xi-linked loci are not informative about the processivity, or the lack thereof, of the DNMT3s. Moreover, when we estimated, with additional constraints in the HMM, the hemi-preference ratio for the DNMT3s from the *FMR1* data (Materials S1), we found that its posterior distribution is essentially the same as its prior (Supplementary Figure 6 in Materials S1), which

indicates that our data are uninformative also about the preference ratio of the DNMT3s.

Behavior of in vivo human methyltransferases at autosomal locus *LEP*

Compared with the two Xi-linked loci, the human *LEP* promoter region is much less methylated and is tissue-specific in the data considered here: the overall methylation density is 18% in the adipose tissues (fat) and 42% in the peripheral blood leukocytes (blood), much lower than 85% at *FMR1* and 82% at *G6PD*. However, most of the methylated CpG dyads are fully methylated: the percentages of fully methylated and hemimethylated dyads are only 16% and 4%, respectively, in fat, and 40% and 5%, respectively, in blood (Table 1). These two *LEP* data sets also contain only a few short runs of hemimethylated dyads (Table 1). Our analysis of these data shows that, although the *LEP* promoter region is sparsely methylated in these tissues, DNMT1 still plays the major role and the rate of maintenance methylation is close to 1 (Table 1; Supplementary Figures 7 and 8 in Materials S1). However, the data are uninformative about what types of process

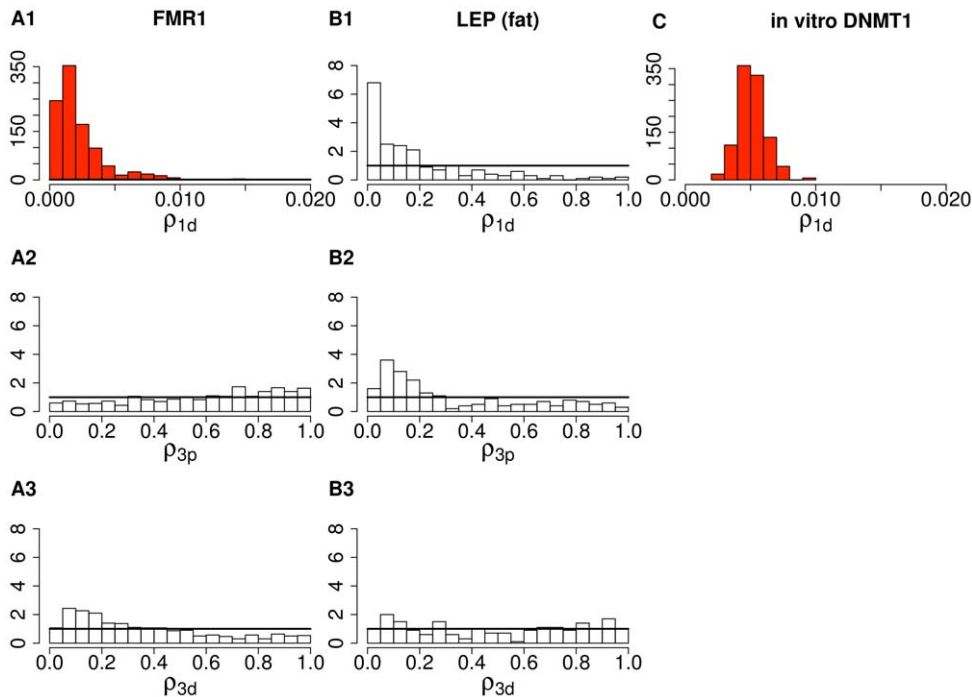


Figure 4. Posterior distributions of dissociating probability ρ from in vivo methylation data at several loci. From top to bottom, each row indicates ρ_{1d} for DNMT1 on the daughter strand, ρ_{3p} for the DNMT3s on the parent strand, and ρ_{3d} for the DNMT3s on the daughter strand. (A1)–(A3) are estimated for the *FMR1* locus. Estimates for *G6PD*, the other Xi-linked locus, show similar distributions and are not displayed here. (B1)–(B3) are estimated for the autosomal *LEP* locus in the fat tissue. Estimates for the same locus in the blood tissue show similar distributions and are not displayed here. (C) is estimated from the data in Goyal et al. [11], one of the four in vitro data sets on mouse DNMT1 we analyze here. Dark lines indicate the density of a uniform (0,1) prior distribution. Except for the plots in red, which have most probability mass around small values, all the other plots are displayed on the same support of (0,1).
doi:10.1371/journal.pone.0032225.g004

(processive or multiple dissociation-reassociation events) gave rise to this high maintenance methylation rate.

Specifically, the *LEP* data are uninformative about the dissociating probability ρ_{1d} and reassociating probability τ_{1d} of DNMT1 in fat and blood, as the posterior distributions of both parameters cover the entire support of (0,1) (Table 2; Figure 4(B1); Supplementary Figures 7 and 8 in Materials S1). Furthermore, we estimated the maintenance activity probability of DNMT1 to be very close to 1 and its de novo activity probability to be very close to 0 (Table 2; also see Supplementary Figures 7 and 8 in Materials S1). Thus, the estimated hemi-preference ratio of DNMT1 at *LEP* is essentially consistent with that at the Xi-linked loci: this ratio is significantly higher than 1, supporting a preference for hemi-methylated CpG sites (Table 2). The median and lower-bound estimates of this ratio at the *LEP* locus is much wider in fat (80% CI: 157–3536) than that in blood, which again suggests that the estimation of this ratio is sensitive to the estimation of the de novo activity probability of DNMT1, and that this ratio can be much higher than the values available in existing literature.

Similar to the *FMR1* and *G6PD* data, the *LEP* data are also uninformative about the DNMT3s on either strand: the posterior distributions of the dissociating probabilities ρ_{3p} and ρ_{3d} are not substantially different from the uniform prior distribution (Table 2; Figure 4(B2)–(B3)). This uninformative nature may have stemmed also from a low level of enzyme activities: the average association frequencies π_{3p} and π_{3d} are not substantially different from 0.

Estimates and impact of measurement errors

Bisulfite conversion used in the experiment can give rise to two types of measurement errors [2,26] (also see Materials and

Methods for details on how we define and incorporate these errors in our analysis). In all the analyses here, we fixed the probability of failure of bisulfite conversion, b , to be 0.003 as in our previous analysis [2]. We estimated the probability of inappropriate bisulfite conversion, c , by taking advantage of the result that this probability has little variation across CpGs in our data set [2]. Our estimates for c are essentially the same in all the data sets, with the posterior median being 2% and narrow 80% CIs (Table 2; also see Supplementary Figure 9 in Materials S1 for the posterior distribution of c estimated for the *FMR1* locus). This appreciable error rate is expected under the low-molarity bisulfite-conversion protocol [26] used to collect our data. Note that the results on processivity and substrate preference given above are robust to different assumptions on the measurement error rates: indeed, setting the error rates to be 0 did not qualitatively impact our inference of the hemi-preference ratio or processivity, except producing a slight reduction in the estimated hemi-preference ratio (Supplementary Figure 10 and Table 1 in Materials S1).

Another source of possible measurement errors is PCR crossover, which can occur during PCR amplification with probability less than 1% per molecule [27], leading to ascertained patterns that are hybrids of two molecules [27]. A crossover between one densely methylated and one sparsely methylated molecule may produce a methylation pattern with one of its ends being mostly hemimethylated dyads, and could affect statistical inference on processivity. Take the *FMR1* data for example. The probability of having at least 1 of 169 patterns produced by a single crossover is 80% under a binomial distribution. Therefore, pattern #82 from this data set (Figure 2), one of the four most informative patterns and the only pattern with a long run of hemimethylated dyads at an end, may

have arisen from a crossover. The other three patterns in Figure 2 have runs of hemimethylated dyads in the middle of the pattern. If these runs were due to crossovers, two events would have had to occur for each molecule. The probability of having at least 1 out of 169 patterns produced by two crossovers is at most 2%. Removing pattern #82 and re-analyzing the rest of the *FMR1* data produced results nearly identical to those for the complete data set (Supplementary Figure 11 in Materials S1). We conclude that PCR crossover errors likely have a negligible impact on our analysis.

Strong processivity of in vitro mouse DNMT1

Most previous studies of DNMT1 were conducted in vitro and investigated mouse DNMT1 [10,11,28,29]. These studies did not consider bisulfite-conversion errors, nor did they distinguish genuine processivity from multiple dissociation-reassociation events at consecutive sites. As explained earlier, these data, containing long runs of as many as 20 hemimethylated dyads, suggest a high level of processivity. We re-analyzed double-stranded methylation patterns from Goyal et al. [11] and Vilkaitis et al. [10] under our HMM, setting the association levels of the DNMT3s to 0, and obtained estimates for the processivity of purified mouse DNMT1 acting in vitro (Supplementary Figure 12 in Materials S1; estimates summarized in Supplementary Table 2 in Materials S1). Consistent with the more descriptive analyses in Goyal et al. [11] and Vilkaitis et al. [10], our statistical analyses also estimated a high level of processivity from these in vitro data on mouse DNMT1, with narrow 80% CIs indicating strong evidence from the data (Figure 5).

Although both in vivo and in vitro data show strong levels of processivity, there are noticeable quantitative differences between the two sets of estimates: our estimates from these published in

vitro data suggest a shorter mean association tract length (Figures 4 and 5A) and a much longer mean non-association tract length (Figure 5B) than do our estimates of human DNMT1 in vivo (see numeric summaries in Table 2 here and Supplementary Table 2 in Materials S1). These disparities could be due to differences between human DNMT1 and mouse DNMT1, as well as to differences among the experimental conditions in these studies. Disparities could also result from differences between in vivo and in vitro conditions, indicating a role for the replication machinery in modulating enzyme activities. Inferences here and elsewhere [10–14] suggest that the replication machinery is not essential for preserving the association of maintenance methyltransferases with the DNA once they are bound. The replication machinery may, however, play a role in keeping unbound DNMT1 poised to re-associate with DNA. This could explain the much shorter non-association tract length and the longer stretches of processive activity inferred here from patterns established in vivo.

Discussion

In this article, we have developed a novel hidden Markov model to infer complex methyltransferase activities from double-strand methylation patterns established in single molecules. This model complements our earlier model [2] that focuses on estimating the CpG site-specific rates of methylation events, regardless of the methyltransferases. Under our HMM, we can estimate reassociating and dissociating probabilities of the enzymes, as well as probabilities of maintenance and de novo activities, inferring association/non-association tract lengths and hemi-preference ratio. Whereas it is possible that the DNMTs may carry out methylation activities in

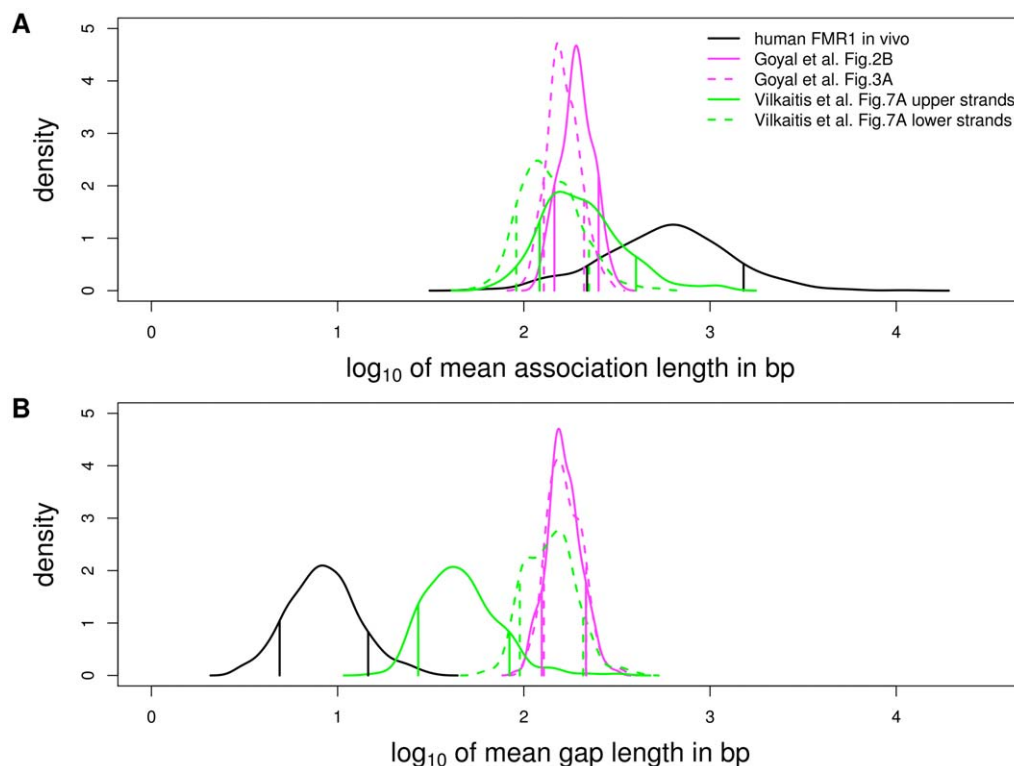


Figure 5. Estimated processivity and non-association tract lengths for human DNMT1 in vivo and mouse DNMT1 in vitro under our HMM. Each curve is the posterior distribution of the (A) mean association (processivity) and (B) non-association tract lengths on the \log_{10} scale. Vertical lines indicate the boundaries of the 80% CIs. Black curves indicate estimates from our in vivo human *FMR1* data. Magenta and green curves are based on our re-analysis of the in vitro mouse DNMT1 data in Goyal et al. [11] and Vilkaitis et al. [10] (Supplementary Figure 12 in Materials S1). doi:10.1371/journal.pone.0032225.g005

multiple passes during a round of DNA replication, the processivity of our HMM infers here is effectively the “cumulative” processivity. It is also unclear how to incorporate multiple passes of methylation with an unknown number of passes into the statistical modeling. Our model is applicable to in vivo data for which possibly both DNMT1 and the DNMT3s were at work. It is also applicable to in vitro data in which only one type of methyltransferase was present. Since the core of our implementation of the inference of our HMM is the standard forward-backward algorithm, the computing time needed for analyzing these data is linearly proportional to the number of patterns and to the number of CpG sites.

Applying our model to four in vivo human data sets collected at three loci, we find strong evidence for a high level of processivity of DNMT1 at two Xi-linked loci, with the mean association tract lengths being a few hundred to a few thousand bp, whereas the methylation patterns at the autosomal locus *LEP* are not informative about processivity. Due to a limited number of loci studied, it is unclear whether the strong processivity of DNMT1 is associated only with the inactive X chromosome. Additionally, the *LEP* locus may not be representative of autosomal loci, because the data were derived from tissue that is composed of different types (adipose tissue contains adipocyte precursors, blood vessels, and stromal cells besides the mature adipocytes). Although the methylation patterns at the *LEP* locus have different densities in the two tissue types, our analysis shows that most of the methylation events at this locus are the maintenance methylation activities due to DNMT1 in both tissue types. At all loci examined here DNMT1 showed a strong preference for hemi-methylated CpG sites over unmethylated sites (point estimates ranged from 15 to 628; obtaining precise estimates is difficult because the denominator of the hemi-preference ratio is the probability of de novo methylation events, and these events are rare in our in vivo data dominated by maintenance methylation events).

Our analysis of in vivo data suggests low contributions from the DNMT3s in these in vivo somatic cells. To study the properties of the DNMT3s, an alternative is to analyze double-stranded methylation patterns from in vitro experiments. Such data are indeed available at least for DNMT3A [30,31]. However, structure analysis suggested that DNMT3A may form a tetramer with DNMT3L, in the form of DNMT3L-DNMT3A-DNMT3A-DNMT3L, which may bind to more than one CpG in a single

binding event [32]. We (AQF and MS) are currently extending our HMM to allow for such possibility and carrying out additional analysis for these DNMT3A data in separate work.

Materials and Methods

Additional details of the Hidden Markov model

Emission probabilities for modeling activities of methyltransferases associated with DNA. Consider the j -th CpG site on the i -th double-stranded methylation pattern. Let Q_{ij} be the methylation state of the post-replication parent CpG at this site and D_{ij} be that of the daughter CpG. Also let M_{ij} , R_{ij}^p and R_{ij}^d be the association or non-association states of the Markov process of DNMT1 at the daughter CpG, and of the DNMT3s at the parent CpG and the daughter CpG, respectively. The emission probabilities, given as $\Pr(Q_{ij}, D_{ij} | M_{ij}, R_{ij}^p, R_{ij}^d)$, are conditional probabilities and computed for the j -th site as in Table 3.

Each entry in Table 3 sums over two states: methylation and no methylation at the pre-replication parent CpG, with probability m_j and $1 - m_j$, respectively. For example, consider the following entry near the bottom right of Table 3, $\Pr((Q_{ij}, D_{ij}) = (1, 1) | (M_{ij}, R_{ij}^p, R_{ij}^d) = (1, 1, 0)) = (1 - m_j)\delta_{1d} + m_j\mu_{1d}$. In this case, DNMT1 and the DNMT3s are both associated with the CpG dyad, with the former at the daughter CpG (i.e., $M_{ij} = 1$) and the latter at the parent CpG (i.e., $R_{ij}^p = 1$, $R_{ij}^d = 0$). Either of two events occurred for the formation of the observed fully methylated CpG dyad (i.e., $(Q_{ij}, D_{ij}) = (1, 1)$): (i) DNMT1 carried out a maintenance methylation event on the daughter CpG of the dyad where the parent CpG had been methylated before replication. This event has probability $m_j\mu_{1d}$; or (ii) the dyad, unmethylated before replication with probability $1 - m_j$, became methylated de novo on the daughter strand by DNMT1 with probability δ_{1d} and on the parent strand by the DNMT3s with probability 1. This double de novo methylation event has total probability $(1 - m_j)\delta_{1d}$.

We made two assumptions in the calculation of the emission probabilities: (i) measurement errors did not occur in the collection of our data; relaxation of this assumption to incorporate error is described below; and (ii) there is no active removal of methyl groups on the parent strand when DNA is replicated [2,21,23,33], which

Table 3. Emission probabilities of the HMM.

(Q_{ij}, D_{ij})				
$(M_{ij}, R_{ij}^p, R_{ij}^d)$	(0,0)	(0,1)	(1,0)	(1,1)
(0,0,0)	$1 - m_j$	0	m_j	0
(0,0,1)	$(1 - m_j)(1 - \delta_{3d})$	$(1 - m_j)\delta_{3d}$	$m_j(1 - \mu_{3d})$	$m_j\mu_{3d}$
(0,1,0)	0	0	1	0
(0,1,1)	0	0	$(1 - m_j)(1 - \delta_{3d}) + m_j(1 - \mu_{3d})$	$(1 - m_j)\delta_{3d} + m_j\mu_{3d}$
(1,0,0)	$(1 - m_j)(1 - \delta_{1d})$	$(1 - m_j)\delta_{1d}$	$m_j(1 - \mu_{1d})$	$m_j\mu_{1d}$
(1,0,1)	$(1 - m_j)(1 - \mu_{3d})(1 - \delta_{3d})$	$(1 - m_j)(\mu_{3d} + \delta_{3d} - \mu_{3d}\delta_{3d})$	$m_j(1 - \mu_{1d})(1 - \delta_{1d})$	$m_j(\mu_{1d} + \delta_{1d} - \mu_{1d}\delta_{1d})$
(1,1,0)	0	0	$(1 - m_j)(1 - \delta_{1d}) + m_j(1 - \mu_{1d})$	$(1 - m_j)\delta_{1d} + m_j\mu_{1d}$
(1,1,1)	0	0	$m_j(1 - \mu_{1d})(1 - \delta_{1d}) + (1 - m_j)(1 - \mu_{3d})(1 - \delta_{3d})$	$m_j(\mu_{1d} + \delta_{1d} - \mu_{1d}\delta_{1d}) + (1 - m_j)(\mu_{3d} + \delta_{3d} - \mu_{3d}\delta_{3d})$

Subscript i denotes the i -th methylation pattern, and j the j -th CpG site. Note that $(M_{ij}, R_{ij}^p, R_{ij}^d)$ denotes the association (1) or non-association (0) state of DNMT1, parent-strand DNMT3s, and daughter-strand DNMT3s, respectively. Also, (Q_{ij}, D_{ij}) denotes the methylated (1) or unmethylated (0) state of CpG on the parent strand and daughter strand, respectively. Additionally, m_j denotes the probability of the j -th CpG site being methylated before DNA replication, which is equivalent to the methylation density of the j -th site. μ_{1d} and μ_{3d} are the probability of the maintenance activity of DNMT1 and the DNMT3s, respectively, at associated daughter-strand CpG, whereas δ_{1d} and δ_{3d} are that of the de novo activity of DNMT1 and the DNMT3s, respectively. The measurement error rates (see Materials and Methods and Table 4) are assumed to be 0 here. This HMM allows for estimation of the hemi-preference ratio for both DNMT1 and the DNMT3s, although additional constraints are needed for this simultaneous estimation. See text for details.

doi:10.1371/journal.pone.0032225.t003

means that a CpG methylated before replication remains methylated after replication. Although active removal of methyl groups has been reported during early development, in cancer cells (see [34] for review), and for transcriptionally active loci under perturbation [35,36], this phenomenon seems uncommon, if it occurs at all, in somatic cells in normal individuals, which are the cell types we study here.

To distinguish between the DNMT1 process and the DNMT3 process on the daughter strand, in the simplest version of the HMM we draw on the existing evidence that the two classes of methyltransferases exhibit different substrate preferences [4,6,7,18,37,38]. When the DNMT3s are associated with the daughter strand, we assume that they add a methyl group with probability 1 ($\mu_{3d} = \delta_{3d} = 1$) at both hemimethylated and unmethylated CpGs. That is, association is synonymous to methylation for the DNMT3s on the daughter strand. In contrast, when DNMT1 is associated with the daughter strand, we allow it to methylate CpGs at hemimethylated sites and at unmethylated sites with different probabilities μ_{1d} and δ_{1d} , respectively, with the ratio μ_{1d}/δ_{1d} being the hemi-preference ratio. To also estimate the hemi-preference ratio for the DNMT3s on the daughter strand, μ_{3d}/δ_{3d} , we use a different set of constraints, namely $\tau_{1d} > 0.05$ and $\tau_{3d} < 0.05$ (Materials S1).

Incorporating measurement errors due to bisulfite conversion. We consider two types of measurement errors due to bisulfite conversion [2,26]: failure of bisulfite conversion, which occurs with probability b per CpG, and inappropriate bisulfite conversion, which occurs with probability c per CpG (see definitions in Table 4). We assume that these errors occur independently across CpGs and DNA strands [2]. Denote Q'_{ij} and D'_{ij} as the observed methylation states at the parent and daughter CpGs, respectively, on the i -th methylation pattern at the j -th CpG dyad, with possible measurement error. Emission probabilities accounting for these measurement errors are:

$$\Pr(Q'_{ij}, D'_{ij} | M_{ij}, R_{ij}^P, R_{ij}^D) = \sum_{Q_{ij}=0,1} \sum_{D_{ij}=0,1} \Pr(Q'_{ij} | Q_{ij}) \Pr(D'_{ij} | D_{ij}) \Pr(Q_{ij}, D_{ij} | M_{ij}, R_{ij}^P, R_{ij}^D), \quad (3)$$

where $\Pr(Q'_{ij} | Q_{ij})$ and $\Pr(D'_{ij} | D_{ij})$ are functions of measurement error probabilities b and c (Table 4), whereas $\Pr(Q_{ij}, D_{ij} | M_{ij}, R_{ij}^P, R_{ij}^D)$ is previously defined as the emission probability without measurement error (Table 3).

Software implementing the HMM. The models and analyses presented here are implemented in the computer program MethylHMM, which can be downloaded from <http://stephenslab.uchicago.edu/software.html>. This program includes also all the data sets analyzed here.

Human in vivo double-stranded methylation data

We used hairpin-bisulfite PCR [21] to collect double-stranded methylation patterns of single molecules from the promoter region of genes *FMR1* and *G6PD*, two loci on the inactive X

Table 4. Probabilities of measurement errors due to bisulfite conversion.

		Observed (Q'_{ij} or D'_{ij})	
		0	1
Truth (Q_{ij} or D_{ij})	0	$1-b$	b
	1	c	$1-c$

True and observed methylation state of the parent-strand CpG in the i -th double-strand methylation pattern at the j -th site are denoted Q_{ij} and Q'_{ij} , respectively. True and observed methylation states on the daughter-strand CpG are denoted D_{ij} and D'_{ij} , respectively. Additionally, b is the probability of failure of bisulfite conversion at a CpG, and c the probability of inappropriate bisulfite conversion.

doi:10.1371/journal.pone.0032225.t004

chromosome, and gene *LEP* on Chromosome 7. Molecular barcodes and batchstamps were used to help identify and remove contaminant and redundant methylation patterns [22]. To capture the variation in methylation patterns across cells, we collected multiple patterns from each individual sampled (Table: 1). Each pattern (see examples in Figure 2) consists of methylation states at CpG sites on the parent and daughter strands at this locus in an individual cell, with no direct information as to which is the parent strand and which is the daughter. The *FMR1* and *G6PD* data were collected on the hypermethylated, inactivated X chromosome in somatic lymphocytes from human females. The *LEP* data were collected from fat tissues (abdomen and breast) and peripheral blood leukocytes. The *FMR1* data were previously analyzed in Fu et al. [2] under a different statistical model that assumed the methylation events at different CpG sites to be independent. The *LEP* data were published in Stöger [39]. Features of the loci and summary statistics of the methylation data are presented in Table 1.

Supporting Information

Materials S1
(PDF)

Acknowledgments

We are grateful to Brooks Miner for contributions to data collection, to Albert Jetsch for the discussion on the structure of DNMT3A binding complex and sharing with us double-stranded methylation data produced in his lab, and to Editor Pierre-Antoine Defossez, two anonymous reviewers and Guillaume Filion for careful reading of our manuscript and constructive comments that have greatly improved our manuscript.

Author Contributions

Conceived and designed the experiments: AQF DPG CDL MS. Analyzed the data: AQF MS. Contributed reagents/materials/analysis tools: AQF. Wrote the paper: AQF DPG CDL MS. Commented on manuscript: RS. Contributed the G6PD data: AFB.

References

- Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 10: 805–811.
- Fu AQ, Genereux DG, Stöger R, Laird CD, Stephens M (2010) Statistical inference of transmission fidelity of DNA methylation patterns over somatic cell divisions in mammals. *Ann Appl Stat* 4: 871–892.
- Bestor TH, Laudano A, Mattaliano R, Ingram VM (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* 203: 971–983.
- Okano M, Xie S, Li E (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* 19: 219–220.
- Kornberg A, Baker T (2005) DNA Replication. University Science Books, second edition. pp 101–112.
- Jeltsch A (2006) On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* 1: 63–66.
- Meselson M, Yuan R (1968) DNA restriction enzyme from *E. coli*. *Nature* 217: 1110–1114.

8. Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71: 865–873.
9. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450: 908–912.
10. Vilkaitis G, Suetake I, Klimasauskas S, Tajima S (2005) Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J Biol Chem* 280: 64–72.
11. Goyal R, Reinhardt R, Jeltsch A (2006) Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic Acids Res* 34: 1182–1188.
12. Schermelleh L, Haemmer A, Spada F, Rösing N, Meilinger D, et al. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res* 35: 4301–4312.
13. Spada F, Haemmer A, Kuch D, Rothbauer U, Schermelleh L, et al. (2007) DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J Cell Biol* 176: 565–571.
14. Feng J, Zhou Y, Campbell SL, Le T, Li E, et al. (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat Neurosci* 13: 423–430.
15. Robertson KD, Wolffe AP (2000) DNA methylation in health and disease. *Nat Rev Genet* 1: 11–19.
16. Hansen RS, Stöger R, Wijmenga C, Stanek AM, Canfield TK, et al. (2000) Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. *Hum Mol Genet* 9: 2575–2587.
17. Li E, Bestor T, Jaenisch R (1996) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69: 915–926.
18. Gowher H, Jeltsch A (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG sites. *J Mol Biol* 309: 1201–1208.
19. Gowher H, Jeltsch A (2002) Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J Biol Chem* 277: 20409–20414.
20. Holz-Schietinger C, Reich N (2010) The inherent processivity of the human de novo methyltransferase 3a (DNMT3A) is enhanced by DNMT3L. *J Biol Chem* 285: 29091–29100.
21. Laird CD, Pleasant ND, Clark AD, Sneed JLS, Hassan KMA, et al. (2004) Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc Natl Acad Sci USA* 101: 204–209.
22. Miner BE, Stöger RJ, Burden AF, Laird CD, Hansen RS (2004) Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. *Nucleic Acids Res* 32: e135.
23. Genereux DP, Miner BE, Bergstrom CT, Laird CD (2005) A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. *Proc Natl Acad Sci USA* 102: 5802–5807.
24. Durbin R, Eddy S, Krogh A, Mitchison G (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids* Cambridge University Press.
25. Liu JS (2004) *Monte Carlo Strategies in Scientific Computing* Springer. pp 105–152.
26. Genereux DP, Johnson WC, Burden AF, Stöger R, Laird CD (2008) Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Res* 36: e150.
27. Burden AF, Manley NC, Clark AD, Gartler SM, Laird CD, et al. (2005) Hemimethylation and non-CpG methylation levels in a promoter region of human LINE-1 (L1) repeated elements. *J Biol Chem* 280(15): 14413–14419.
28. Bestor TH, Ingram VM (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc Natl Acad Sci USA* 80: 5559–5563.
29. Hermann A, Goyal R, Jeltsch A (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* 279: 48350–48359.
30. Jurkowska RZ, Anspach N, Urbanke C, Jia D, Reinhardt R, et al. (2008) Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Res* 36: 6656–6663.
31. Rajavelu A, Jurkowska RZ, Fritz J, Jeltsch A (2011) Function and disruption of DNA Methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. doi:10.1093/nar/gkr753.
32. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3l suggests a model for de novo DNA methylation. *Nature* 449: 248–251.
33. Fu AQ (2008) *Models and Inference of Transmission of DNA Methylation Patterns in Mammalian Somatic Cells*. Ph.D. thesis, University of Washington.
34. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11: 204–220.
35. Métivier R, Gallais R, Tiffiche C, Le Péron C, Jurkowska RZ, et al. (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452: 45–52.
36. Kangaspek S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, et al. (2008) Transient cyclical methylation of promoter DNA. *Nature* 452: 112–116.
37. Arber W (1974) DNA modification and restriction. *Prog Nucleic Acid Res Mol Biol* 23: 185–188.
38. Smith SS, Hardy TA, Baker DJ (1987) Human DNA (cytosine-5)methyltransferase selectively methylates duplex DNA containing mispairs. *Nucleic Acids Res* 15: 6899–6916.
39. Stöger R (2006) In vivo methylation patterns of the leptin promoter in human and mouse. *Epigenetics* 1: 155–162.